Information intersections: interaction and diffusion of knowledge between academic literature, the social web, and the mass media / Hadas Shema

Abstract

This dissertation presents three articles, all studying scholarly blogs and their potential as sources of scholarly impact. The first article, "Research blogs and the discussion of scholarly information," was published on May 2012 in the journal PLOS ONE. The second article, "Do blog citations correlate with a higher number of future citations? Research blogs as a potential source for alternative metrics," has been published on May 2014 in the Journal of the Association for Information Science and Technology (JASIST) and the third article, "How is research blogged? A content analysis approach," has been published on advanced online publication in JASIST as well.

All three articles studied blogs aggregated in ResearchBlogging.org (RB), a blog aggregator for blog posts referring to peer-reviewed works. Using a code snippet, RB bloggers can cite peer-reviewed works in a structured, formal manner, similar to the reference format used in peer-reviewed publications.

In the first article we characterized blogs and bloggers in order to shed light on this form of scholarly discourse. The sample consisted of non-commercial blogs that aggregated at least 20 items with RB between January 1st, 2010 and January 15th, 2011 and had 1-2 authors. We found that RB bloggers often cited articles from top-tier multidisciplinary and biomedical journals. The most popular journals were (in alphabetical order): Nature, PLOS ONE, PNAS and Science. The blogs mainly focused on the life and behavioral sciences. Seventy-two percent of the blogs in the sample had at least one active, public Twitter account associated with them and at least 90% of these accounts followed or were followed by another Twitter account from the sample. A majority of the bloggers were men (76% of the blogs had at least one male author), about a third (32%) had been awarded a PhD, 27% were graduate students, 59% were affiliated with a research institute and most (84%) blogged under their own name.

The first article showed that most of the bloggers who chose to formally cite the articles that they blogged about were highly educated and affiliated with a research institute. We hypothesized that, given the bloggers' scholarly background, they might be able to

recognize and cover articles in their blog which will later become better cited compared with articles from the same journal and year which were not covered in blogs. To test this hypothesis, we collected RB data from blog posts published during the years 2009 and 2010 and the articles they cited, limiting the sample to journals that had at least 20 articles covered in a blog during 2009 or 2010 (some journals were included in the sample for 2009 and 2010, but some made it only to the 2009 or 2010 sample). Most journals that passed the 20-article threshold were prestigious multidisciplinary and biomedical journals, similarly to the first study's results. Nature, PLOS ONE, PNAS and Science were again the most covered journals. We collected citation data about the blog-covered articles and about their journals in general from Web of Science (WoS) from 2009, 2010 and 2010 for the 2009 articles and from 2010, 2011 and 2012 for the year 2010. We ran a non-parametric one-sample binomial test and found that for the entire sample in the year 2009 and in the year 2010 the proportion of blogged articles that received above their journals' general median counts for the same year was more than 0.5. We also ran non-parametric Mann-Whitney tests to test whether individual journals' articles covered in blogs had a citation advantage over articles from the same year that were not covered in blogs and found that the results were statistically significant for 7 out of 12 journals (58%) in 2009 and 13 out of 19 journals (68%) in 2010. Based on these results, we suggested that scholarly blogs might be a promising source of alternative metrics for scholarly impact.

In the third article we studied the apparent motivations behind RB blog posts using content analysis methods. The sample consisted of blog posts from the years 2010-2012. We sampled 10% of RB's health category at random, overall analysing 391 posts. Similarly to the first and second studies, Nature, PNAS, PLOS ONE and Science were the most covered journals, followed by general medicine journals (e.g., the Lancet). We created a classification scheme with ten major categories (discussion, criticism, advice, trigger, extensions, self, controversy, data, ethics and other) that each had several subcategories. The vast majority of the blog posts (about 90%) included a general discussion of the issue covered in the article, almost 30% of the posts included some criticism of the issues being discussed (not necessarily of the article cited). Over a quarter of the posts offered advice of some sort, showing that the bloggers were willing to share their knowledge and expertise. The relatively high percentage of criticism (compared with studies of traditional citations) suggests that perhaps the informal style of blogs allows for the easier expression of negativity. While the first article merely described scholarly blogs citing peer-reviewed research in a formal manner, the second suggested

they might be a useful source of scholarly impact metrics and the third attempted to reveal, using content analysis methods, the motivations behind individual blog posts.

System No.

002392410